IN THE SPECIFICATION

Please amend the paragraph at page 23, line 3 to line 12 as follows.

In one embodiment, the substrings are randomly concatenated to produce "recombined" strings. In one approach to such "random" concatenation, each substring is assigned a unique identifier (*e.g.* an integer or other identifier). The identifiers are then randomly selected from the pool (*e.g.* using a random number generator) and the subsequences corresponding to those identifiers are joined to produce a concatenated sequence. When joined subsequences are approximately [ore] <u>or</u> exactly the length of the stating character string(s), the process is started anew to produce another string. The process is repeated until all of the substrings are utilized. Alternatively the substrings can be selected without withdrawing them from the "substring pool" and the process is repeated until a desired number of "full-length" strings are obtained.

Please amend the paragraph at page 21, line 1 to line 19 as follows.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information [(http://www.nebi.nlm.nih.gov/)]. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W,T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix *(see*

2

E a coli.

Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.